

Applied A.I. Solutions

Data Visualization Techniques

Professor
Daniel Vitaver, EMBA

daniel.vitaver@georgebrown.ca

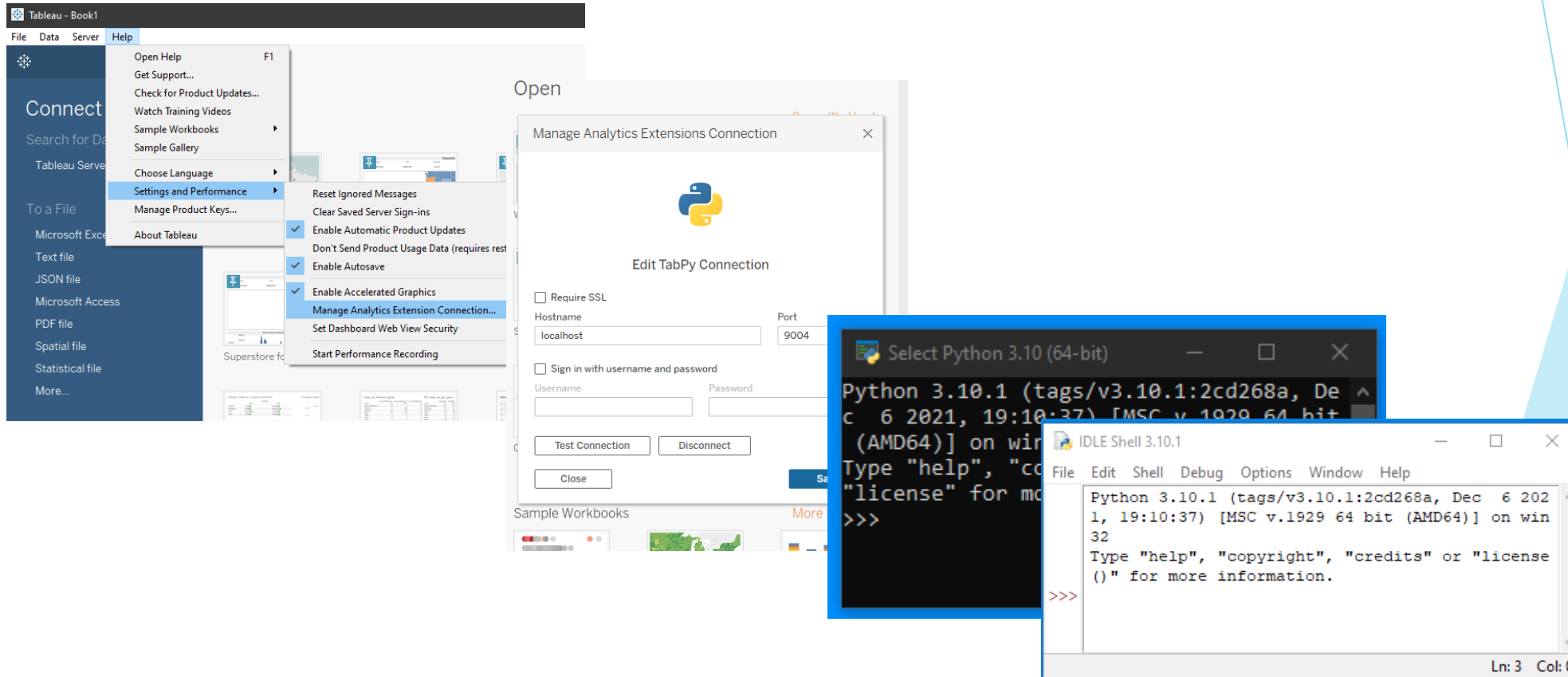
DATA VISUALIZATION TECHNIQUES

Tableau - Python Integration

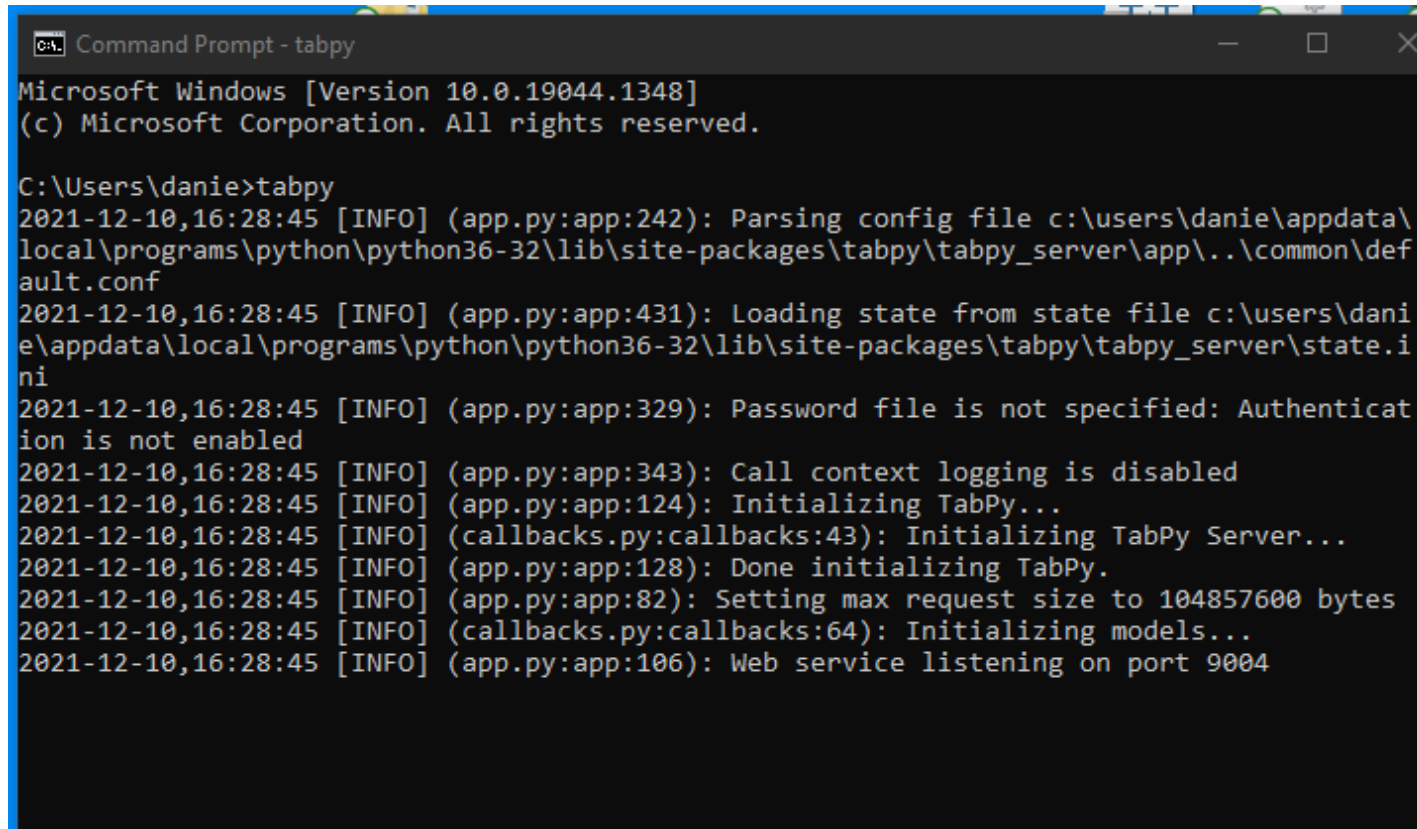
Tableau Python Server - TabPy

- TabPy is an API that enables evaluation of Python code from within a Tableau workbook.
- When you use TabPy with Tableau, you can define calculated fields in Python, leveraging the power of machine-learning libraries right from your visualizations.
- This Python integration in Tableau enables powerful scenarios. For example, **it takes only a few lines of Python code to get data you can explore in using Tableau.**

1. Open Tableau Pro Desktop
2. Test Python/TabPy Connection to Tableau (localhost, 9004)
3. Run Python
4. Run IDLE (Integrated Development and Learning Environment)



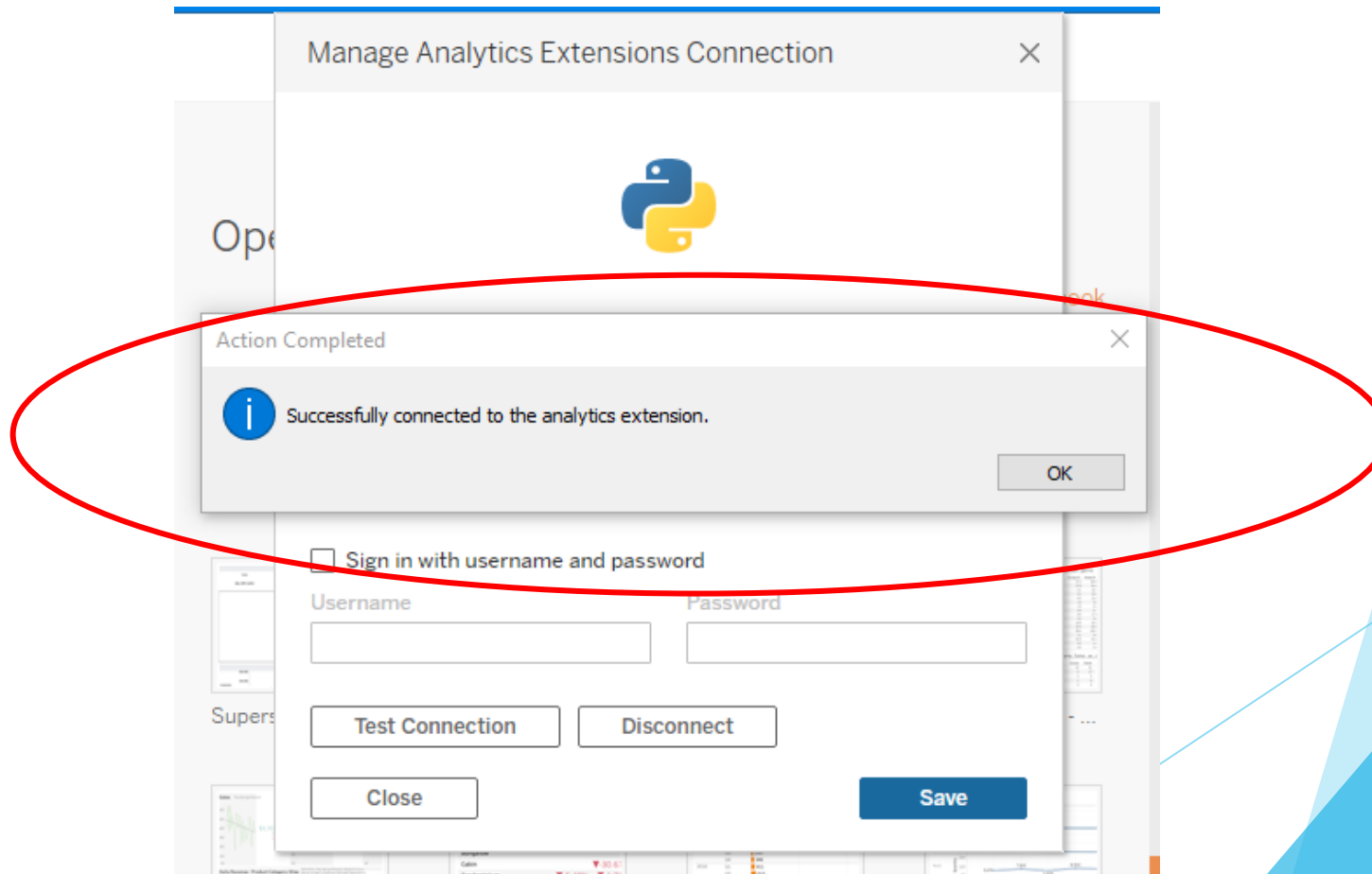
5. Run `c:\` Command Prompt
6. From `c:\` run *`pip install TabPy`* (*install TabPy*)
7. From `c:\` run *`TabPy`*

A screenshot of a Windows Command Prompt window titled "Command Prompt - tabpy". The window shows the output of running the "tabpy" command from the C:\ directory. The output consists of several lines of log messages from the TabPy application, including parsing the config file, loading the state file, and initializing the server and models. The messages are timestamped and include file paths and log levels (INFO).

```
Microsoft Windows [Version 10.0.19044.1348]
(c) Microsoft Corporation. All rights reserved.

C:\Users\danie>tabpy
2021-12-10,16:28:45 [INFO] (app.py:app:242): Parsing config file c:\users\danie\appdata\
local\programs\python\python36-32\lib\site-packages\tabpy\tabpy_server\app\..\common\def
ault.conf
2021-12-10,16:28:45 [INFO] (app.py:app:431): Loading state from state file c:\users\dani
e\appdata\local\programs\python\python36-32\lib\site-packages\tabpy\tabpy_server\state.i
ni
2021-12-10,16:28:45 [INFO] (app.py:app:329): Password file is not specified: Authenticat
ion is not enabled
2021-12-10,16:28:45 [INFO] (app.py:app:343): Call context logging is disabled
2021-12-10,16:28:45 [INFO] (app.py:app:124): Initializing TabPy...
2021-12-10,16:28:45 [INFO] (callbacks.py:callbacks:43): Initializing TabPy Server...
2021-12-10,16:28:45 [INFO] (app.py:app:128): Done initializing TabPy.
2021-12-10,16:28:45 [INFO] (app.py:app:82): Setting max request size to 104857600 bytes
2021-12-10,16:28:45 [INFO] (callbacks.py:callbacks:64): Initializing models...
2021-12-10,16:28:45 [INFO] (app.py:app:106): Web service listening on port 9004
```

5. Leave all the opened environments running
6. Go to Tableau Desktop
7. Test Python/TabPy Connection to Tableau (localhost, 9004)
8. Press OK
9. Save



To verify the installation of TabPy in a web browser type:

<http://localhost:9004/>

The browser renders



Now you are ready to open Tableau Desktop

[GitHub - tableau/TabPy](#)

[TabPy Installation Instructions | TabPy \(tableau.github.io\)](#)

Importing Libraries in Tableau calculated fields

Importing Libraries in Tableau calculated fields

- `import pandas as pd`
- `import numpy as np`
- `import matplotlib.pyplot as plt`
- `import seaborn as sns`
- `%matplotlib inline`
- `import warnings`
- `warnings.filterwarnings('ignore')`
- `from sklearn.preprocessing import LabelEncoder`

Python Functions

- To let Tableau, know that the calculations are using TabPy, it must be passed through one of the four python functions
- Tableau aggregated measures like **SUM(), ATTR(), MIN(), MAX()** can **also be included** in TabPy calculations
- TabPy functions are **always** computed as Table calculations in Tableau

Python Functions

Python functions that are used in the calculated field to indicate TabPy are:

- **SCRIPT_REAL**: returns an output of type real from the given calculation
- **SCRIPT_INT**: returns an output of type integer from the given calculation
- **SCRIPT_STR**: returns an output of type string from the given calculation
- **SCRIPT_BOOL**: returns an output of type Boolean from the given calculation

Whenever any of these functions are used in a calculated field the calculation is passed directly to a **running TabPy external service instance**.

TabPy Example for each Python function

SCRIPT_BOOL: Profitable Sales

1. We use Sample **Superstore** data to find the Profitable Sales of products
2. Since the profitable Sales can be either true or false, we use **SCRIPT_BOOL**
3. Tableau aggregates data **before** sending it to TabPy using the level of detail of the view.

Python – Numpy library Integration

1

TabPy - 1

Results are computed along Table (across).

```
SCRIPT_REAL("
ratio = []
for i in range(0, len(_arg1)):
    ratio.append(_arg2[i] / _arg1[i] * 100)
return ratio
", SUM([Sales]), SUM([Profit]))
```

2

TabPy - 2

Results are computed along Table (across).

```
SCRIPT_REAL("
import numpy as np
p = np.polyld([0.5, 100])
return p(_arg1).tolist()
", SUM([Sales]))
```

3

Profit Ratio

```
sum([Profit]) / sum([Sales])
```

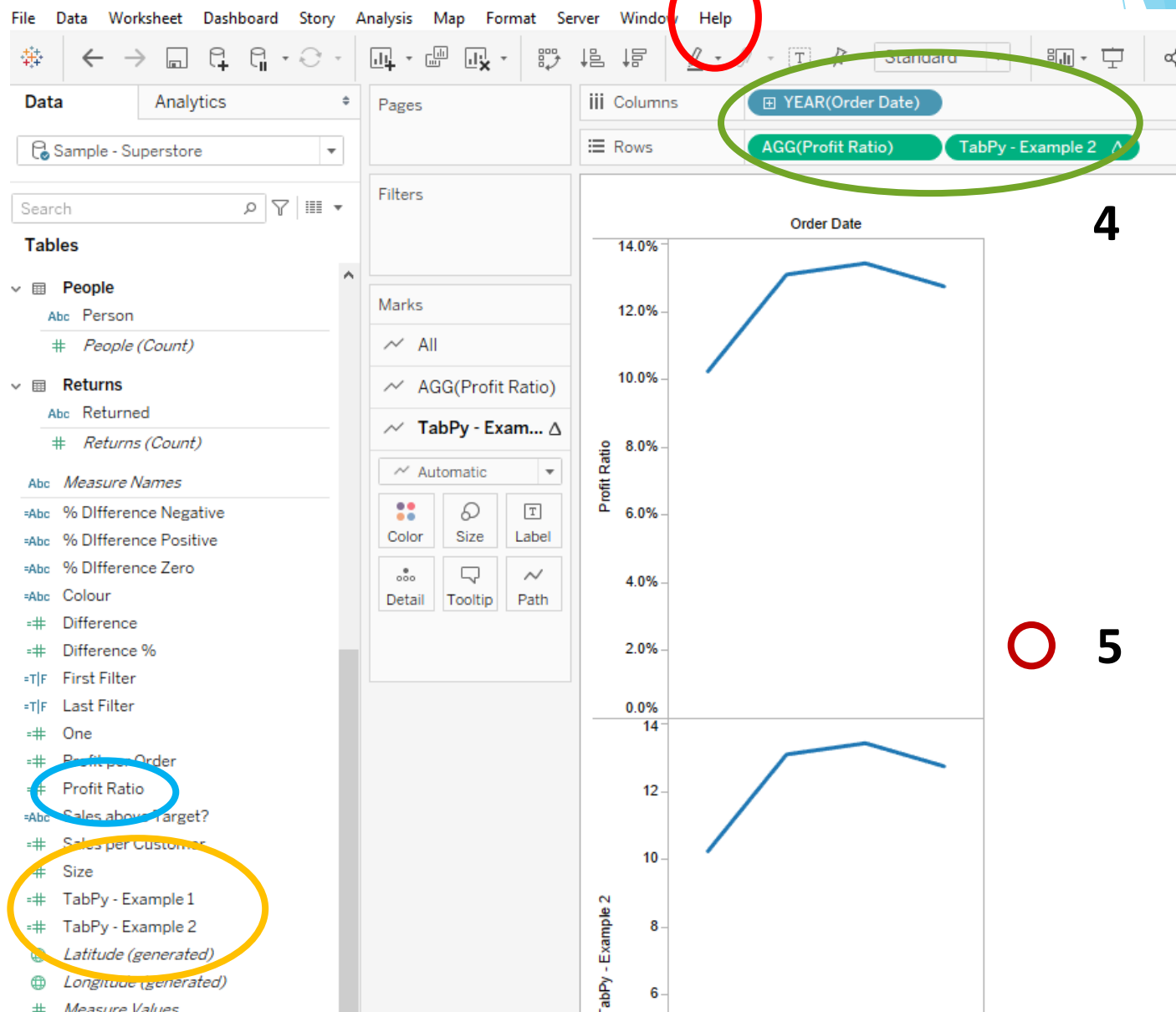
The calculation is valid. 17 Dependencies

Apply OK

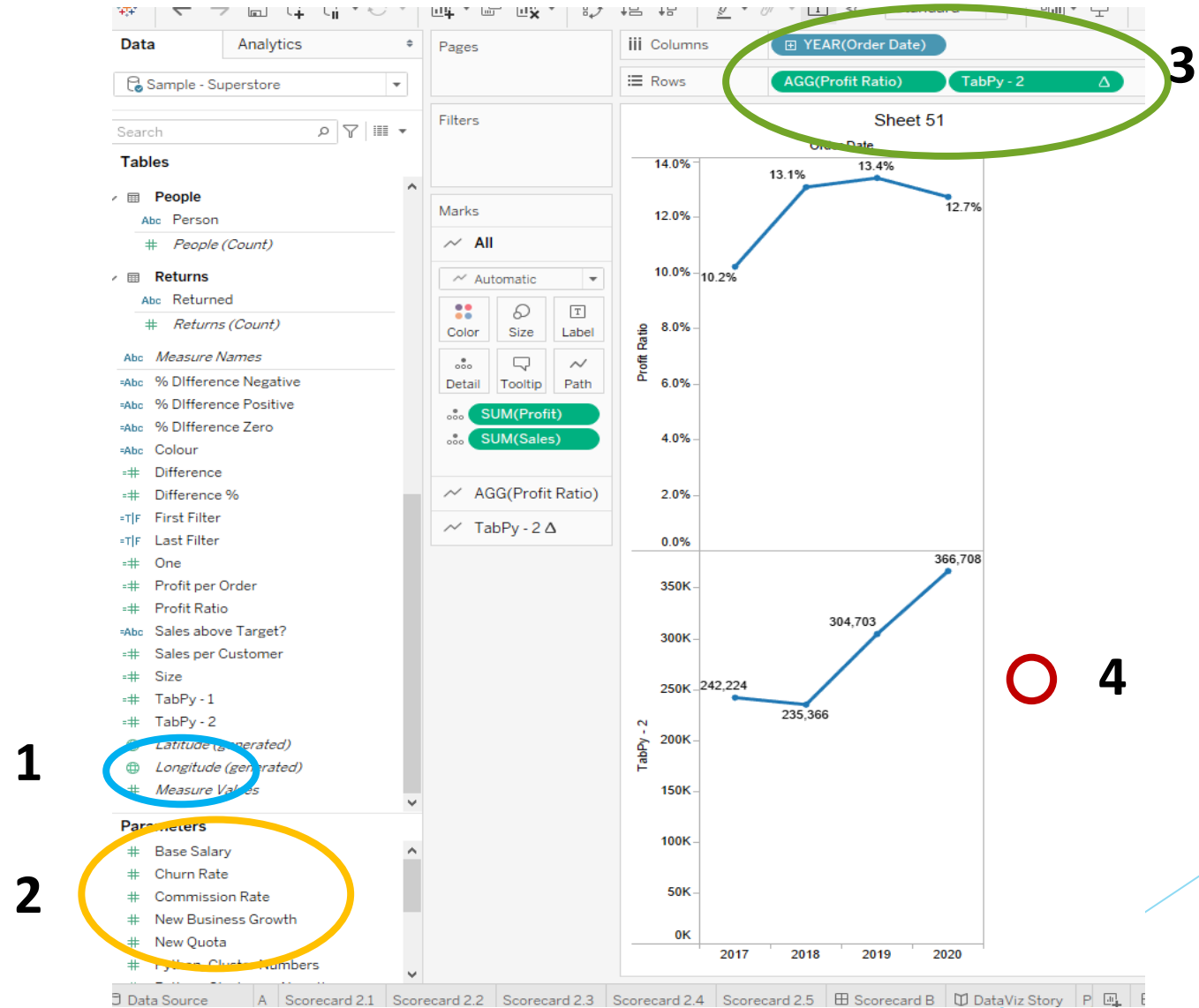
Python – Numpy library Integration

3

2



Python – Numpy library Integration



TabPy Example 1

Python calculation of Profitable Sales

Profitable Sales

Sample - Superstore

X

Results are computed along Table (across).

```
SCRIPT_BOOL("  
lst = []  
for i in _arg1:  
    lst.append(i>0)  
return lst  
",  
SUM([Profit])  
)
```

Default Table Calculation

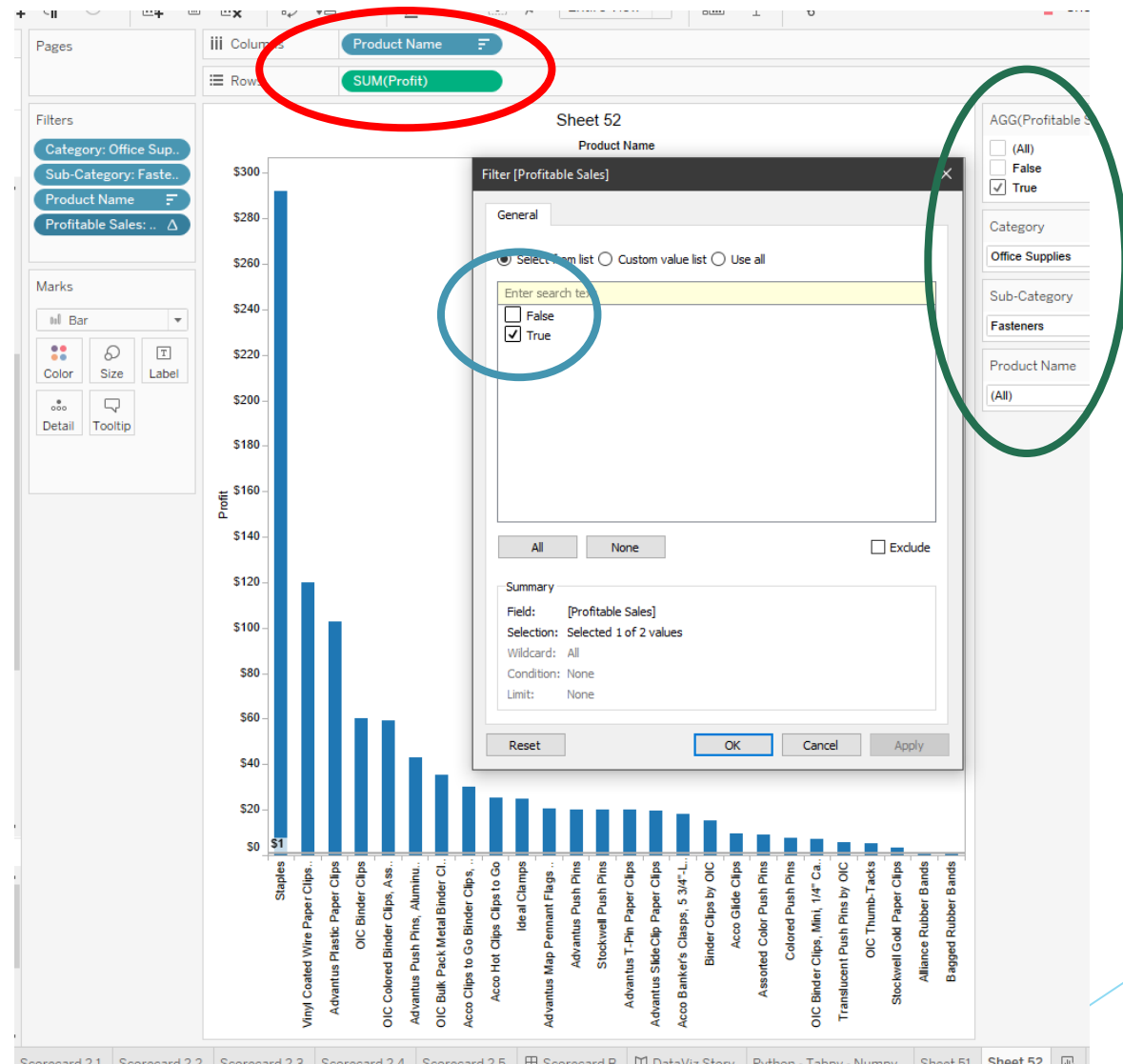
The calculation is valid.

Apply

OK

See
importing
Python
libraries to
Tableau

Python calculation of Profitable Sales



TabPy Example 2

Pearson Correlation Coefficient

- The Pearson coefficient is a type of correlation coefficient that represents the **relationship between two variables that are measured on the same interval or ratio scale.**
- The Pearson coefficient is a measure of the **strength** of the association between two continuous variables
- Pearson correlation coefficient or the bivariate correlation is the measure used to measure the **linear relationship between two variables.**
- Correlation can either be **positive or negative** indicating if a particular variable is impacting another variable positively or negatively.
- **Pearson's correlation should be used only when there is a linear relationship between variables.** It can be a positive or negative relationship, as long as it is significant.

Python calculation of Profitable Sales

2 Pearson Correlation Sample - Superstore X

Results are computed along Customer Name.

```
SCRIPT_REAL("import numpy as np
return np.corrcoef(_arg1,_arg2)[0,1]",
SUM([Sales]),SUM([Profit]))|
```

1

See importing Python libraries to Tableau

Table Calculation [Pearson Correlation] X

Calculation Definition

Compute using:

3 Customer Name

At the level:

Restarting every:

Description

Results are computed along Customer Name.

OK

Cancel

The calculation is valid.

1 Dependency

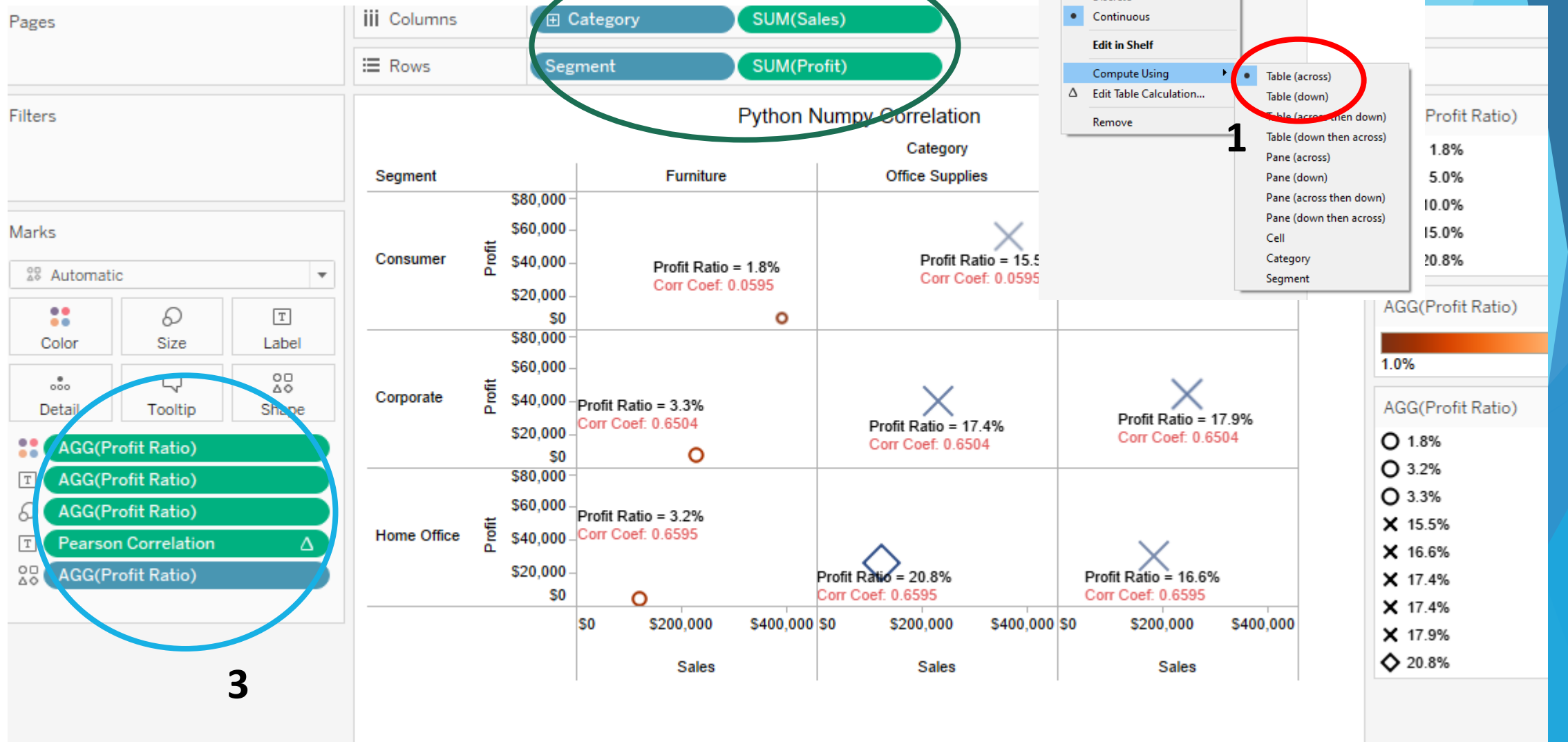
Default Table Calculation

Apply

4 OK

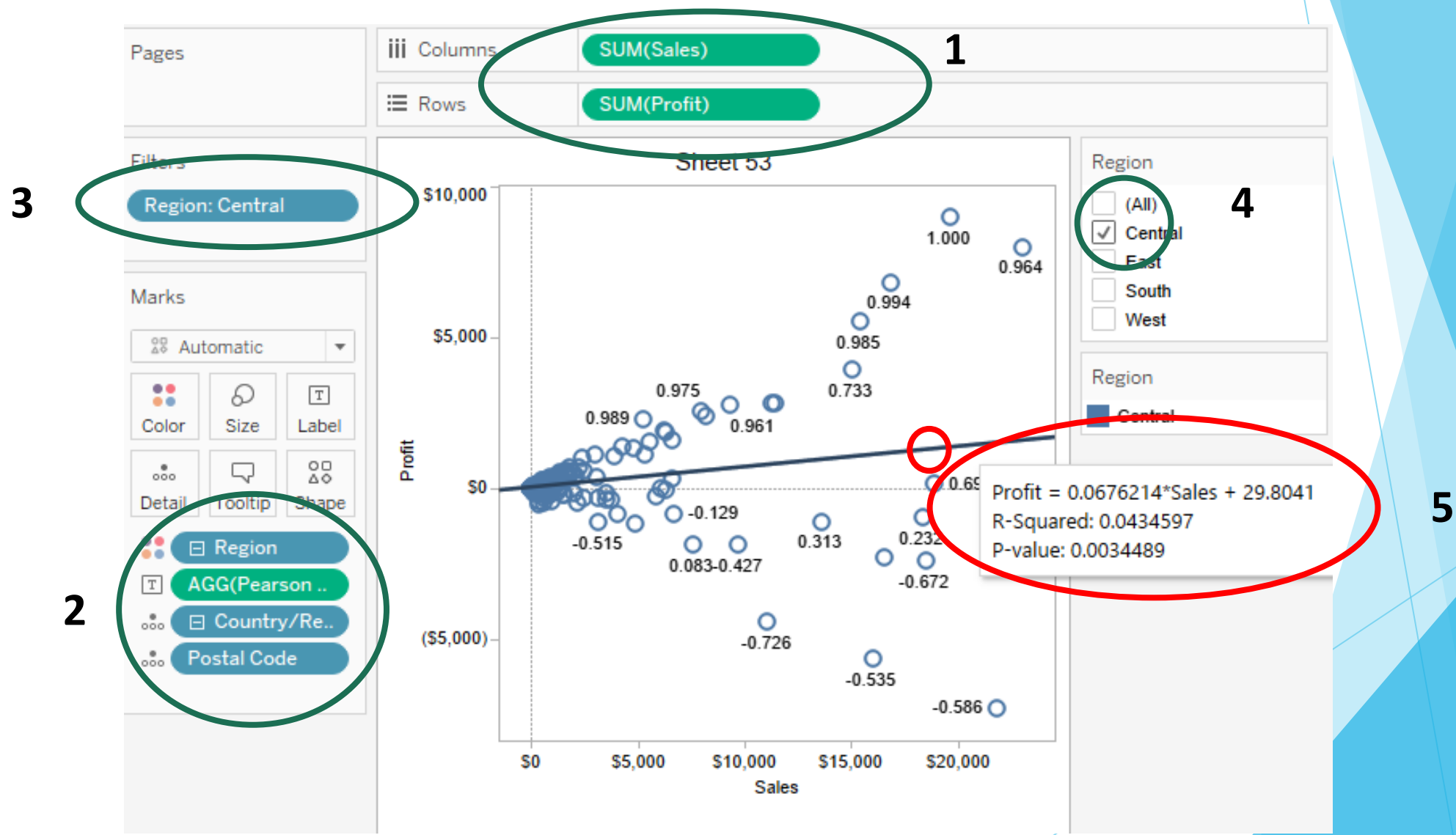
Pearson Correlation Calculation

2



3

Correlation Using Tableau Only – Function CORR

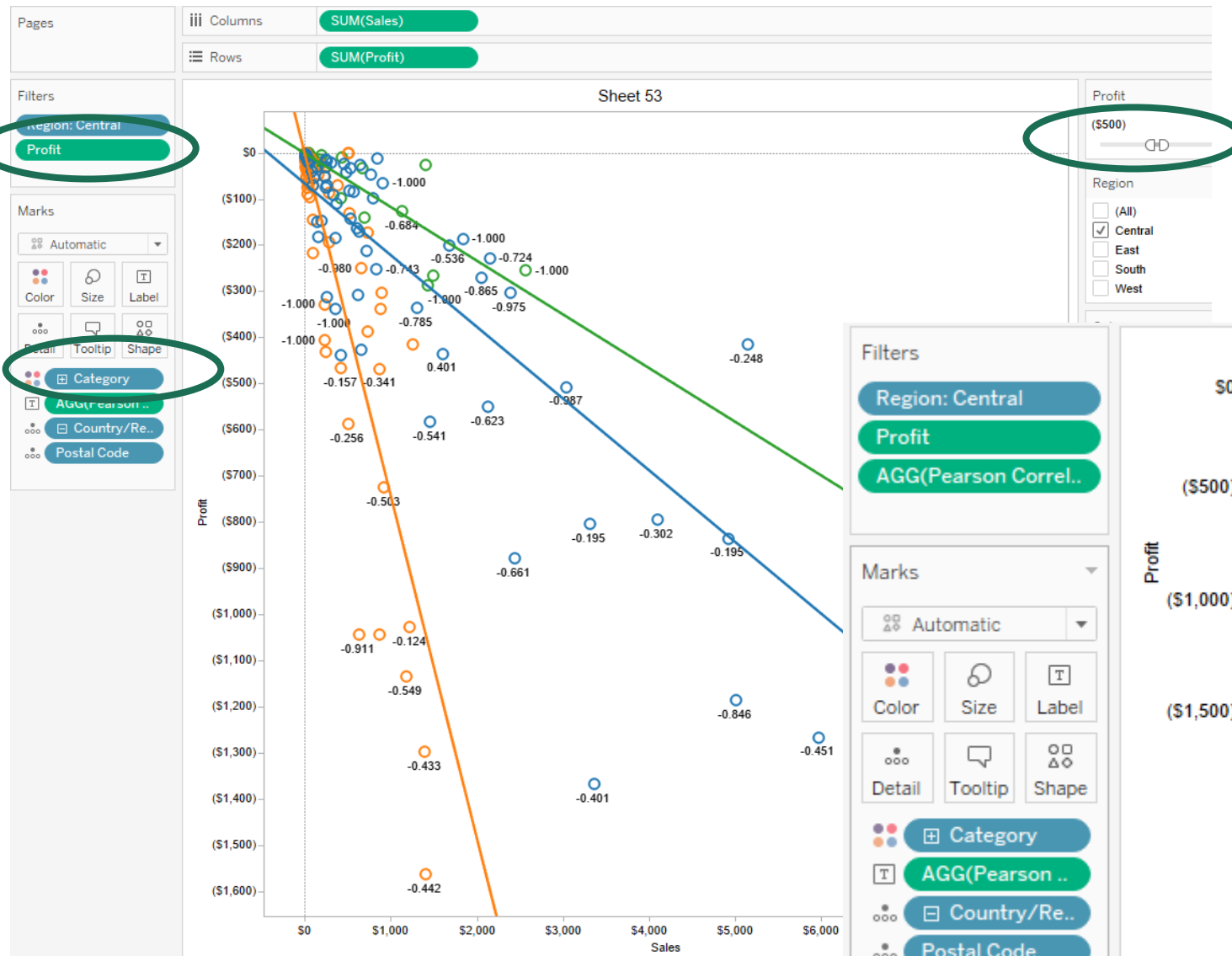


1

2

3

4



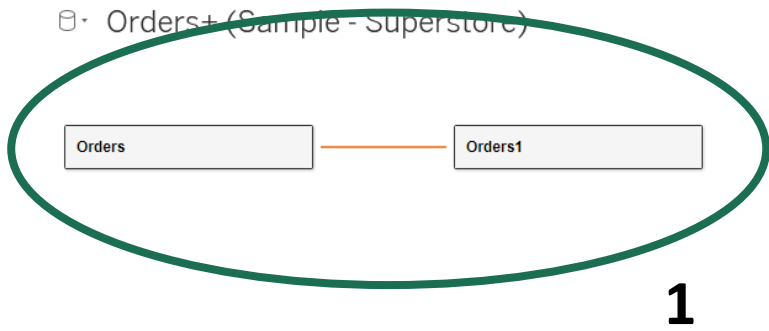


Correlation Matrix Using Tableau Only – Function CORR

<https://www.youtube.com/watch?v=vhrmmvAvOfQ>

[Tableau Custom Charts series: Download Supplemental Materials - Page - SuperDataScience | Machine Learning | AI | Data Science Career | Analytics | Success](#)

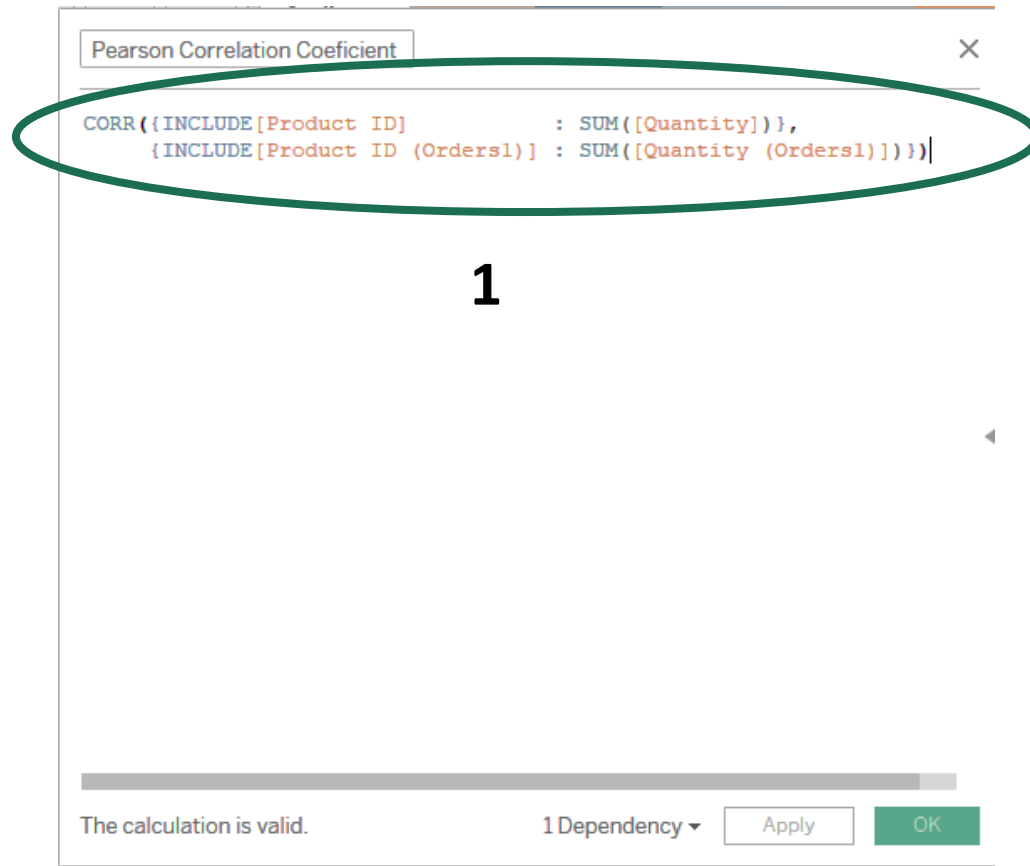
Correlation Matrix Using Tableau Only – Function CORR

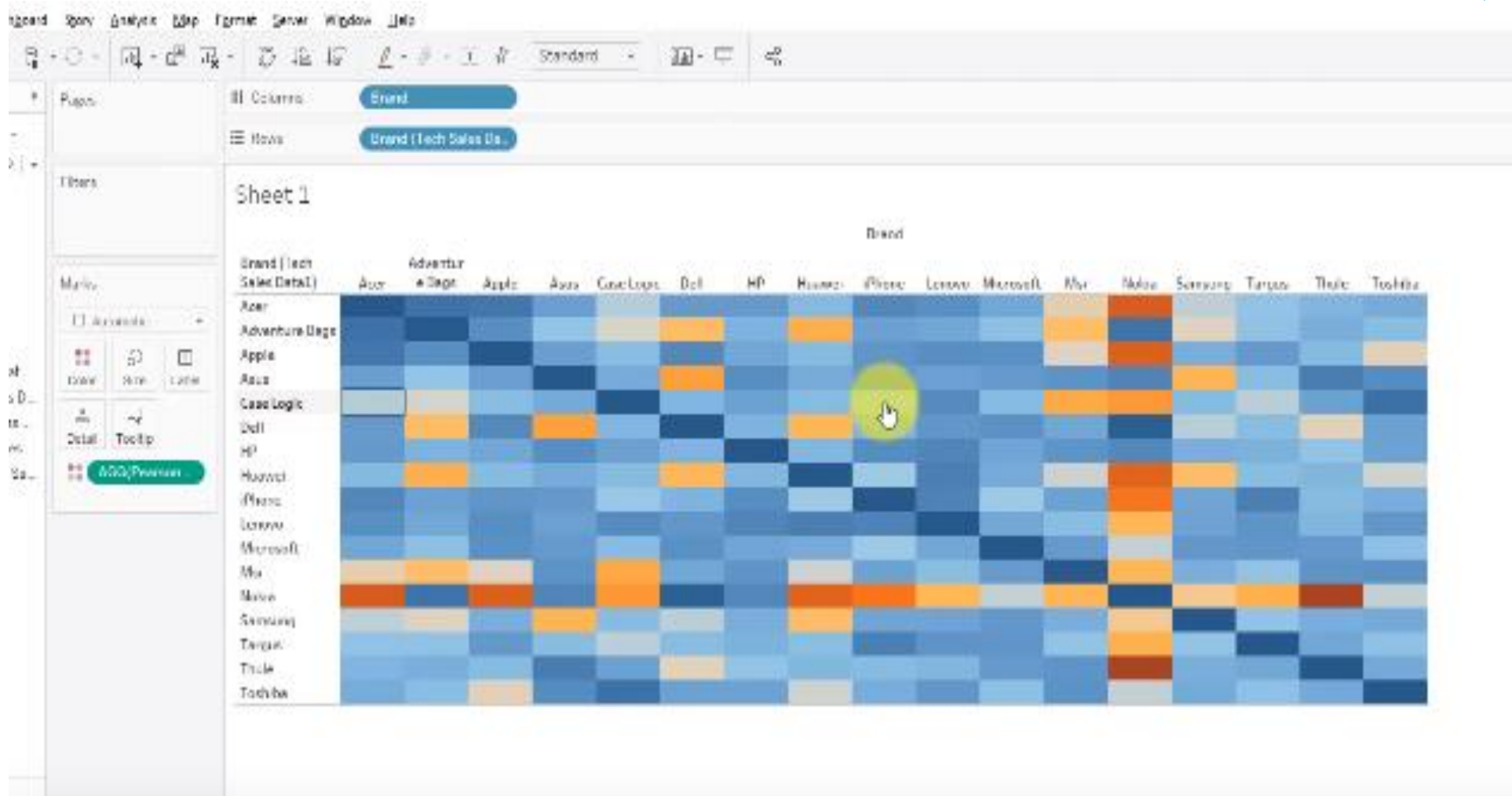


Edit Relationship			
How do relationships differ from joins? Learn more			
Orders		Orders1	
Order ID		=	Order ID (Orders1)
Add more fields			
Performance Options			
Standard Class			

#	Order ID	Order Date	Ship Date
1	CA-2019-152156	11/8/2019	11/11/2019
2	CA-2019-152156	11/8/2019	11/11/2019
3	CA-2019-138688	6/12/2019	6/16/2019
4	US-2019-100955	10/11/2019	10/18/2019

Correlation Matrix Using Tableau Only – Function CORR





Correlation: must watch videos

- **Tableau Correlation Matrix**

<https://www.youtube.com/watch?v=vhrmmvAvOfQ>

- **How to find the Pearson correlation in Tableau**

<https://www.youtube.com/watch?v=Ns8HulbYnp8>

- **NumPy, SciPy, and Pandas: Correlation With Python**

<https://realpython.com/numpy-scipy-pandas-correlation-python/#pearson-correlation-coefficient>

